# Does systematic sampling preserve granger causality with an application to high frequency financial data?

Rajaguru, Gulasekaran; Abeysinghe, Tilak; O'Neill, Michael

# 2017 Australian Conference of Economists, Sydney, 19-21 July 2017

## Does Systematic Sampling Preserve Granger Causality with an Application to High Frequency Financial Data?

Tilak Abeysinghe, Michael O'Neill and Gulasekaran Rajaguru

# Motivation

*There was a lot of high powered analysis of this topic, but I came away from a reading of it with the feeling that it was one of the most unfortunate turnings for econometrics in the last two decades, and it has probably generated more nonsense results than anything else during that time.*

-Pagan (1989)

# Motivation

- Financial Development (FD) and Economic Growth (EG)
  - Controversial causal results
    - **FD → EG:** McKinnon (1973), King and Levine (1993a,b), Neusser and Kugler (1998), and Levine et al. (2000), Christopoulos and Tsionas (2004), Dimitris and Efthymios (2004), Shan (2006), and Zhang, Wang and Wang (2012)
    - **EG → FD**: Gurley and Shaw (1967), Goldsmith (1969), Jung (1986) and Liang and Teng (2006)
    - **Bidirectional:** Shan, Morris, and Sun (2001) and Demetriades and Hussein (1996), Calderón and Liu (2003), Hassan, Fung (2009), Sanchez and Yu – (2011)and Kar et al. (2011)

## Literature Review

**Temporal Aggregation / Systematic Sampling**

Univariate ARIMA models: Wei(1990)

Dynamic relationships: Temporal Aggregation weakens the distributed lag relationships, Telser(1967), Zellner and Montmarquette(1971), Sims(1971), Wei and Tiao (1975), Tiao and Wei(1976,1978), Wei and Metha (1980)

**Causality: Wei (1982), Mamingi (1996), Marcellino (1999), Rajaguru (2004), Rajaguru and Abeysinghe ( 2008), Abeysinghe and Rajaguru (2012) and Ashley and Tsang(2014)**

Forecasting: Lutkepohl (1987), Marcellino (1999)

Unit Roots: Pierce and Snell(1995)
Co-integration: Phillips(1991)

Misspecification involved in cross-country regression: Ericsson et al,(2000)

Temporal aggregation turns one-way causality into feedback system

Systematic sampling preserves the causal directions

# Systematic Sampling

- Let $z_t = (z_{1t}, z_{2t}, \dots, z_{nt})$, $t=1,2,\dots,T$ be an equally spaced n-variate basic disaggregated series.
- Systematic sampling : $Z_\tau = z_{m\tau}$ ($\tau =1,2,\dots,N$ and $T=mN$) - sampling from $z_t$ at every $m^{th}$ interval ($m$ is a positive integer).

# Systematic Sampling

- Let $w_t = (w_{1t}, w_{2t}, \ldots, w_{nt})$, $w_{jt} = (1-L)^{d_j} z_{jt}$, be a weakly stationary process.
- where $\gamma^w{}_{ii}(k)$ is the autocovariance of the $i$-th component, $w_{it}$ at lag k.
- $\gamma^w{}_{ij}(k)$ is the cross covariance between $i$-th and $j$-th components
- $\gamma^w{}_{ii}(0)$ is the variance of the $i$-th series

# Relationship between disaggregated and Systematic Sampled Series

$$\mathrm{W}_{\mathrm{j}\tau} = (1-L')^{d_j} Z_{j\tau} = (1-L^m)^{d_j} z_{jm\tau} = (1+L+.....+L^{m-1})^{d_j} w_{jm\tau} \,.$$

- The $d_j$-*th* difference of the systematically sampled series ($j$-th component) is simply the weighted sum of the $d_j$-*th* difference of the basic series.

# Relationship between cross-covariances of disaggregated and Systematic Sampled Series

**Proposition 1**

The cross covariance between *i*-th and *j*-th components of the systematically sampled series $W_{i\tau}$ and $W_{j\tau-k}$ can be expressed in terms of cross covariances of the *i*-th and *j*-th components of the basic disaggregated series $w_{it}$ and $w_{jt}$, that is,

$$\gamma_{ij}^{W}(k) = Cov(W_{i\tau}, W_{j\tau-k}) = (1 + L + L^2 + \ldots\ldots + L^{m-1})^{d_i + d_j} \gamma_{ij}^{w}(mk + d_j(m-1)) \qquad (1)$$

$$\gamma_{ji}^{W}(k) = Cov(W_{j\tau}, W_{i\tau-k}) = (1 + L + L^2 + \ldots\ldots + L^{m-1})^{d_i + d_j} \gamma_{ji}^{w}(mk + d_i(m-1)) \qquad (2)$$

# Systematic Sampling and Granger Causality

consider the following bivariate VAR(1) system with $z_{1t} \sim I(d_1)$ and $z_{2t} \sim I(d_2)$ such that $w_{it} = (1-L)^{d_i} z_{it}$ for *i=1,2*:

$$\begin{pmatrix} w_{1t} \\ w_{2t} \end{pmatrix} = \begin{pmatrix} \varphi_{11} & \varphi_{12} \\ \varphi_{21} & \varphi_{22} \end{pmatrix} \begin{pmatrix} w_{1t-1} \\ w_{2t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}, \quad \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} \sim N\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \right),$$

- $\varphi_{12} \neq 0$ implying Granger causality from $w_2$ to $w_1$
- $\varphi_{21} \neq 0$ implying Granger causality from $w_1$ to $w_2$

# Systematic Sampling and Granger Causality

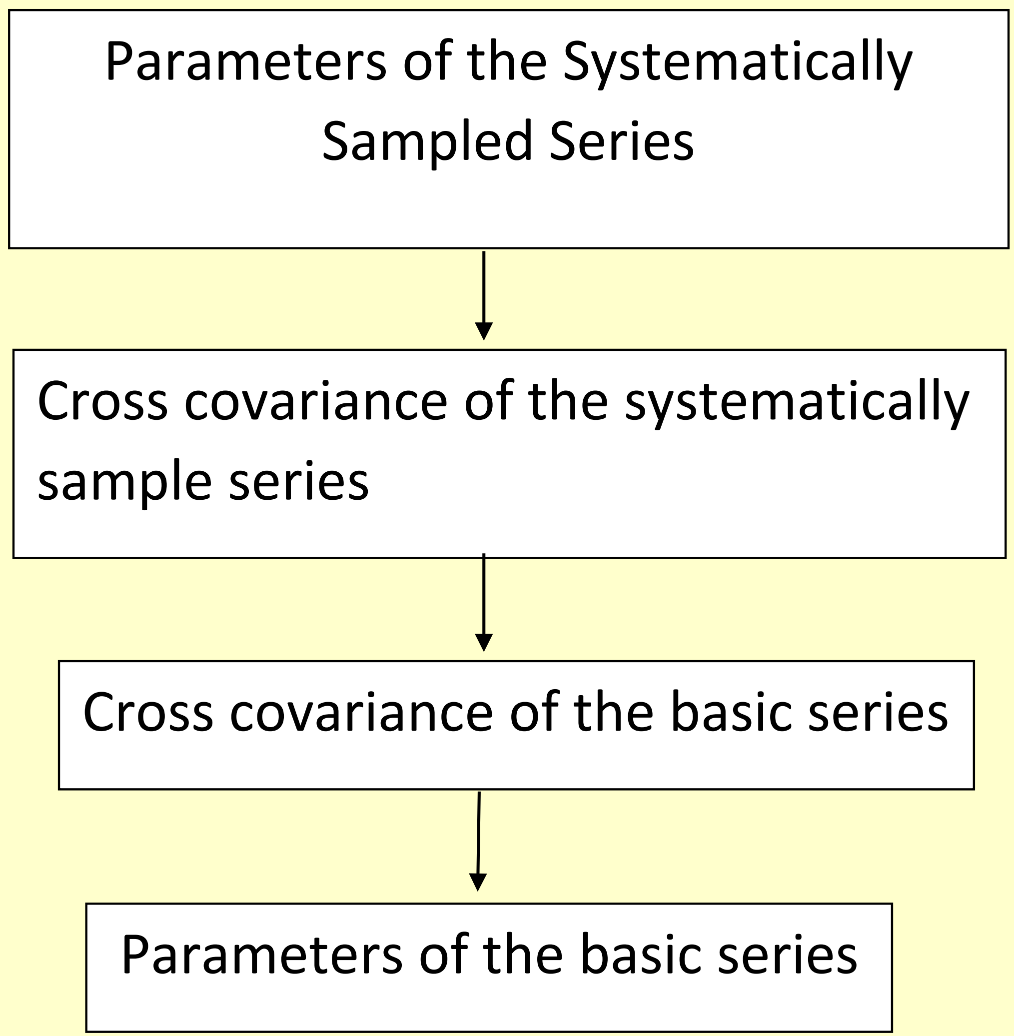
consider the following bivariate VAR(1) system based on systematically sampled series:

$$\begin{pmatrix} W_{1\tau} \\ W_{2\tau} \end{pmatrix} = \begin{pmatrix} \varphi_{11}^{*} & \varphi_{12}^{*} \\ \varphi_{21}^{*} & \varphi_{22}^{*} \end{pmatrix} \begin{pmatrix} W_{1\tau-1} \\ W_{2\tau-1} \end{pmatrix} + \begin{pmatrix} E_{1\tau} \\ E_{2\tau} \end{pmatrix},$$

$$p\lim \hat{\phi}_{11}^{*} = \frac{\gamma_{11}^{W}(1)\gamma_{22}^{W}(0) - \gamma_{12}^{W}(1)\gamma_{12}^{W}(0)}{\gamma_{11}^{W}(0)\gamma_{22}^{W}(0) - \left(\gamma_{12}^{W}(0)\right)^{2}},\ p\lim \hat{\phi}_{12}^{*} = \frac{\gamma_{12}^{W}(1)\gamma_{11}^{W}(0) - \gamma_{11}^{W}(1)\gamma_{12}^{W}(0)}{\gamma_{11}^{W}(0)\gamma_{22}^{W}(0) - \left(\gamma_{12}^{W}(0)\right)^{2}},$$

$$p\lim \hat{\varphi}_{21}^{*} = \frac{\gamma_{21}^{W}(1)\gamma_{22}^{W}(0) - \gamma_{22}^{W}(1)\gamma_{12}^{W}(0)}{\gamma_{11}^{W}(0)\gamma_{22}^{W}(0) - \left(\gamma_{12}^{W}(0)\right)^{2}},\ p\lim \hat{\varphi}_{22}^{*} = \frac{\gamma_{22}^{W}(1)\gamma_{11}^{W}(0) - \gamma_{21}^{W}(1)\gamma_{12}^{W}(0)}{\gamma_{11}^{W}(0)\gamma_{22}^{W}(0) - \left(\gamma_{12}^{W}(0)\right)^{2}}$$

# Systematic Sampling and Granger Causality



# Case 1: No Granger causality between the variables in the disaggregated form

$$\begin{pmatrix} w_{1t} \\ w_{2t} \end{pmatrix} = \begin{pmatrix} \varphi_{11} & \varphi_{12} \\ \varphi_{21} & \varphi_{22} \end{pmatrix} \begin{pmatrix} w_{1t-1} \\ w_{2t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}$$

Here $\varphi_{12} = \varphi_{21} = 0$ and with $\sigma_{12} = 0$

**Proposition 2**

*If there does not exist Granger causality between the basic series then the Granger causality between the systematically sampled series is also absent.*

# Case 2: Causality between the disaggregated series is one-sided

$$\begin{pmatrix} w_{1t} \\ w_{2t} \end{pmatrix} = \begin{pmatrix} \varphi_{11} & \varphi_{12} \\ \varphi_{21} & \varphi_{22} \end{pmatrix} \begin{pmatrix} w_{1t-1} \\ w_{2t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}$$

Here $\phi_{12} = 0$ and $\phi_{21} \neq 0$ and with $\sigma_{12} = 0$

**Theorem 1**

*Systematic sampling induces spurious bi-directional Granger causality among the variables if the uni-directional causality runs from a non-stationary series to either a stationary or a non-stationary series.*

*Equivalently, systematic sampling induces spurious bi-directional Granger causality among the variables if* $d_1 > 0$.

# *Case 3: Causality between the disaggregated series is bi-directional*

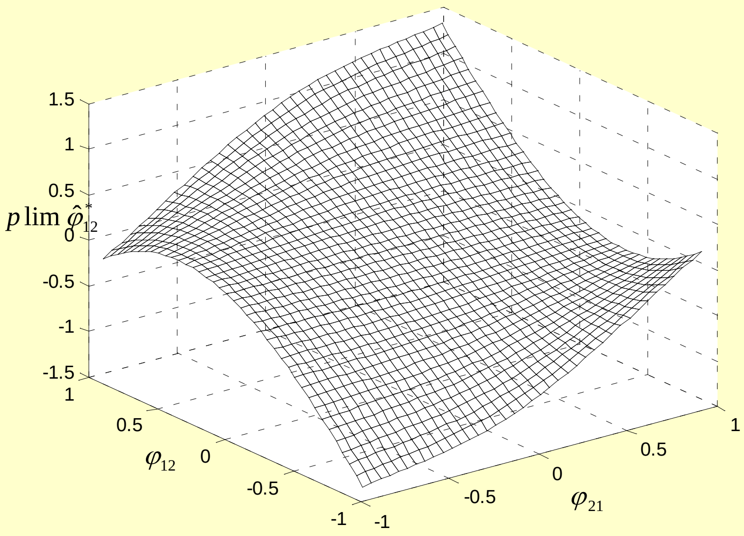
$$\begin{pmatrix} w_{1t} \\ w_{2t} \end{pmatrix} = \begin{pmatrix} \varphi_{11} & \varphi_{12} \\ \varphi_{21} & \varphi_{22} \end{pmatrix} \begin{pmatrix} w_{1t-1} \\ w_{2t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}$$

Here $\phi_{12} \neq 0$ and $\phi_{21} \neq 0$ and with $\sigma_{12} = 0$

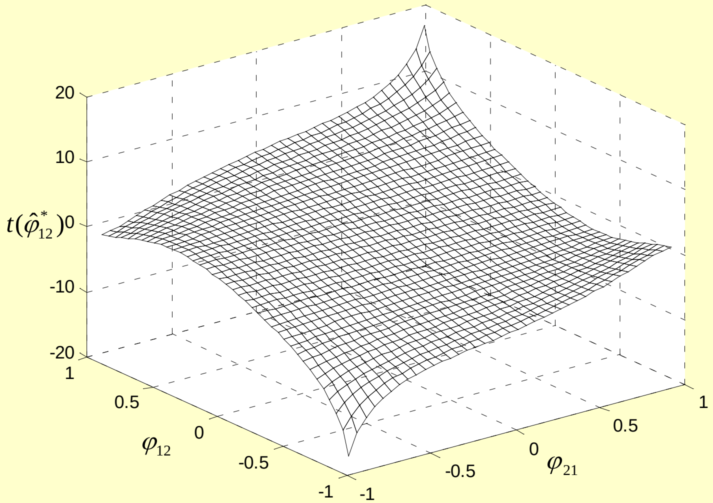
- *Bi-directional causal system becomes uni-directional at the lower level of aggregation (systematic sampling) and subsequently becomes no-causality among the variables of interest*
- *All causal informations concentrate on contemporaneous relationships at the higher level of aggregation*

# *Case 3: Monte Carlo Simulation*

$p\lim \hat{\varphi}_{12}^{*}$ from a feedback system when $\varphi_{11} = \varphi_{22} = 0$, ***m=3*** and $d_1 = d_2 = 0$



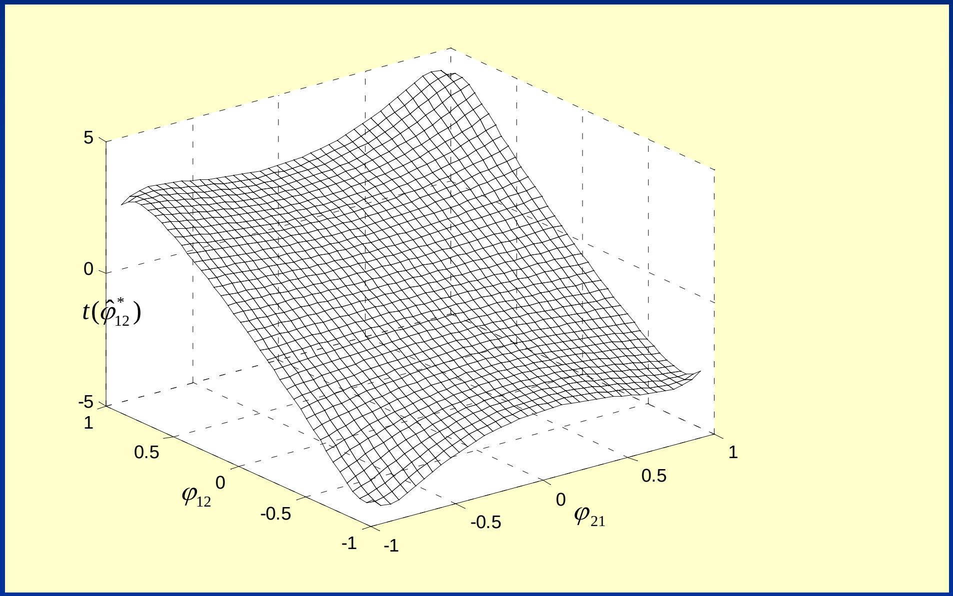

m=3



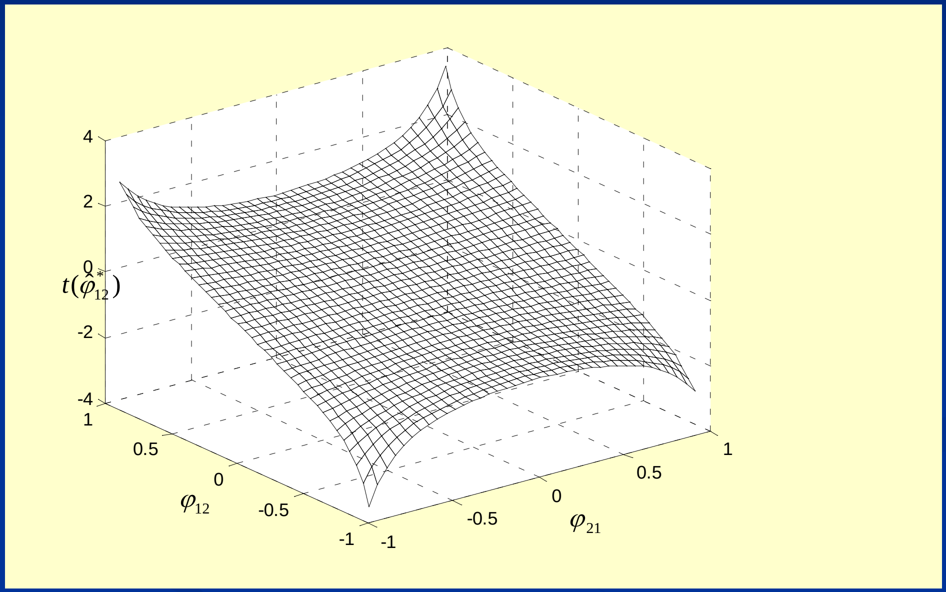

m=3

# Case 3: Monte Carlo Simulation

$t(\hat{\phi}^*_{12})$ from a feedback system when $\varphi_{11} = \varphi_{22} = 0$, ***m=12 and 60*** and $d_1 = d_2 = 0$



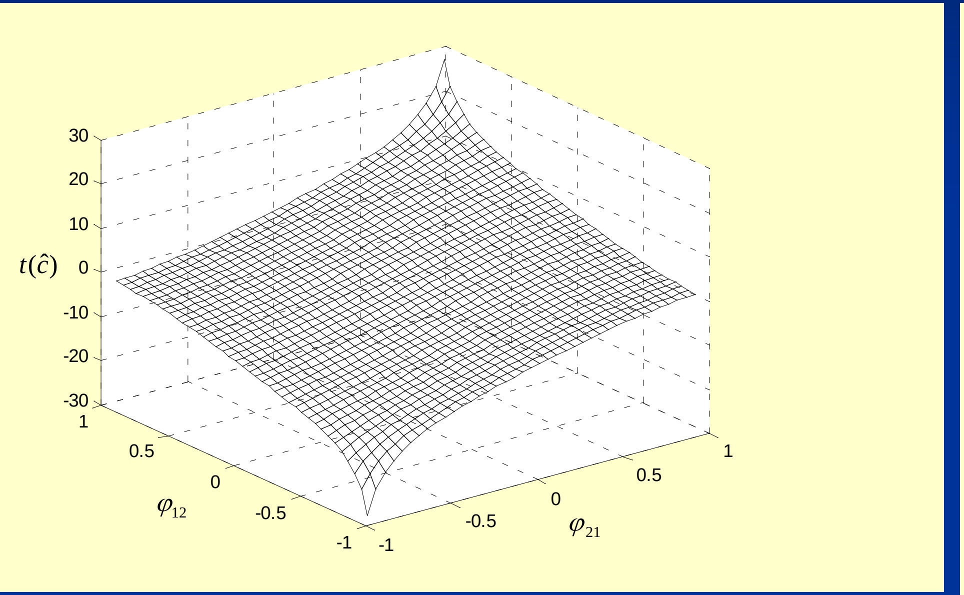

m=12



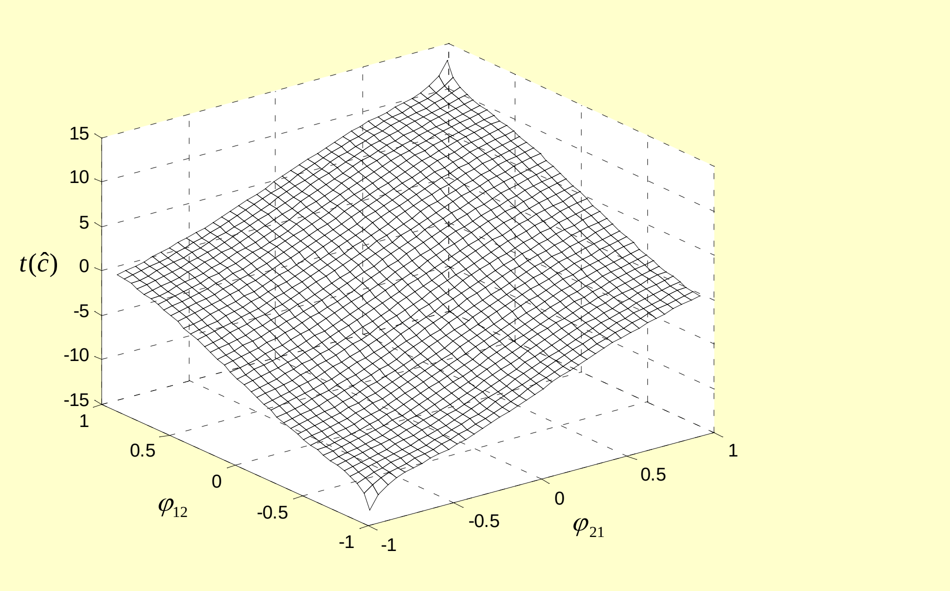

m=60

# *Case 3: Monte Carlo Simulation*

$t(\hat{c})$ from a feedback system when $\varphi_{11} = \varphi_{22} = 0$, ***m=12 and 60*** and $d_1 = d_2 = 0$




m=12




m=60

Contemporaneous Regression

$$w_{1t} = cw_{2t} + v$$

# Applications

- Bi-directional Granger causality has been highlighted in studies using high frequency data (Frijns et al, 2015; Bollen, O’Neill and Whaley, 2016).
- Data sourced from Jan 2010 to Dec 2014 from Thompson Reuters SIRCA portal and Bloomberg:

  -US equities: S&P 500 index (SPX).

  -Equity index futures: E-Mini futures index (SC1/ES1).

  -“Investor fear gauge”: CBOE Volatility Index (VIX).

  -Futures on VIX: S&P VIX futures short term index (SPVXSTR/VST)

# Application 1: *SPX vs VIX I(0)/I(0)*

$$\begin{pmatrix} SPX_t \\ VIX_t \end{pmatrix} = \begin{pmatrix} \varphi_{11} & \varphi_{12} \\ \varphi_{21} & \varphi_{22} \end{pmatrix} \begin{pmatrix} SPX_{t-1} \\ VIX_{t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}$$

| | | 1-minute | | | | 5-minutes | | | | 10-minutes | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **Both** | **SPX** | **VIX** | **None** | **Both** | **SPX** | **VIX** | **None** | **Both** | **SPX** | **VIX** | **None** |
| **15 Sec** | **Both** | 63 | 38 | 20 | 3 | 42 | 33 | 16 | 33 | 26 | 12 | 4 | 82 |
| | **SPX** | 81 | 590 | 14 | 96 | 40 | 388 | 8 | 345 | 22 | 107 | 0 | 652 |
| | **VIX** | 11 | 12 | 59 | 15 | 5 | 5 | 34 | 53 | 0 | 1 | 9 | 87 |
| | **None** | 7 | 19 | 21 | 200 | 2 | 11 | 10 | 224 | 0 | 3 | 4 | 240 |

Note: Rejection frequencies of Granger non-causality at the 5% level of significance

# Application 2: *VIX vs SPVXSTR I(0)/I(1)*

$$\begin{pmatrix} VIX_t \\ \Delta VST_t \end{pmatrix} = \begin{pmatrix} \varphi_{11} & \varphi_{12} \\ \varphi_{21} & \varphi_{22} \end{pmatrix} \begin{pmatrix} VIX_{t-1} \\ \Delta VST_{t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}$$

| | | 1-minute | | | | 5-minutes | | | | 10-minutes | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **Both** | **VIX** | **VST** | **None** | **Both** | **VIX** | **VST** | **None** | **Both** | **VIX** | **VST** | **None** |
| **15 Sec** | **Both** | 566 | 118 | 322 | 23 | 32 | 176 | 335 | 486 | 14 | 226 | 143 | 646 |
| | **VIX** | 2 | 57 | 1 | 2 | 0 | 49 | 0 | 13 | 0 | 35 | 0 | 27 |
| | **VST** | 9 | 21 | 7 | 2 | 4 | 23 | 4 | 8 | 1 | 17 | 2 | 19 |
| | **None** | 1 | 3 | 2 | 16 | 0 | 2 | 0 | 20 | 0 | 1 | 0 | 21 |

Note: Rejection frequencies of Granger non-causality at the 5% level of significance

# Application 3: *ES1 vs* SPVXSTR *I(1)/I(1)*

$$\begin{pmatrix} \Delta ES1_t \\ \Delta VST_t \end{pmatrix} = \begin{pmatrix} \varphi_{11} & \varphi_{12} \\ \varphi_{21} & \varphi_{22} \end{pmatrix} \begin{pmatrix} \Delta ES1_{t-1} \\ \Delta VST_{t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}$$

| | | 5-minutes | | | | 10-minutes | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Both | ES1 | VST | NONE | Both | ES1 | VST | NONE |
| 1-minute | Both | 219 | 383 | 237 | 147 | 63 | 176 | 198 | 549 |
| | ES1 | 8 | 54 | 6 | 26 | 5 | 21 | 7 | 61 |
| | VST | 7 | 6 | 19 | 14 | 1 | 2 | 11 | 32 |
| | NONE | 1 | 0 | 5 | 20 | 0 | 0 | 1 | 25 |

Note: Rejection frequencies of Granger non-causality at the 5% level of significance

# Conclusion

- If there does not exist Granger causality between the basic series then the Granger causality between the systematically sampled series is also absent.
- Systematic sampling induces spurious bi-directional Granger causality among the variables if the uni-directional causality runs from a non-stationary series to either a stationary or a non-stationary series.
- As m increases VAR(1) becomes VAR(0).
  - All causal inferences concentrate on contemporaneous relationship among the variables due to systematic sampling of integrated process. However, interestingly, the spurious contemporaneous relationships do not disappear even if the sampling interval is larger.

Thank you!